

STAT626_MRI

Philip Anderson; panders2@tamu.edu

6/3/2018

```
library("tidyverse")
library("astsa")
```

Objective:

I am doing a freestyle EDA process to explore the data and get myself and everyone else more acquainted with it. I'll focus on the techniques we have covered so far.

Import the data

```
# read in the data
mri <- read.csv("/Users/panders2/Documents/schools/tamu/stat_626/project/stat_626_proj/mri_dat_one.csv")
# update fields
names(mri) <- c("hour", "minute", "freq", "int_pressure", "atm_pressure", "tot_pressure", "tesla")
# take a look
str(mri)
```

```
## 'data.frame': 1603 obs. of 7 variables:
## $ hour : int 0 0 0 0 0 0 0 0 0 0 ...
## $ minute : int 1 2 3 4 5 6 7 8 9 10 ...
## $ freq : num 63800687 63800687 63800687 63800687 63800688 ...
## $ int_pressure: num 2.95 2.95 2.96 2.97 2.97 ...
## $ atm_pressure: num 14.7 14.7 14.7 14.7 14.7 ...
## $ tot_pressure: num 17.6 17.6 17.6 17.6 17.7 ...
## $ tesla : num 1.5 1.5 1.5 1.5 1.5 ...
```

```
head(mri)
```

```
##   hour minute      freq int_pressure atm_pressure tot_pressure tesla
## 1    0      1 63800687         2.950         14.682         17.632 1.49851
## 2    0      2 63800687         2.953         14.682         17.635 1.49851
## 3    0      3 63800687         2.963         14.682         17.645 1.49851
## 4    0      4 63800687         2.967         14.682         17.649 1.49851
## 5    0      5 63800688         2.974         14.682         17.656 1.49851
## 6    0      6 63800688         2.979         14.682         17.661 1.49851
```

```
summary(mri)
```

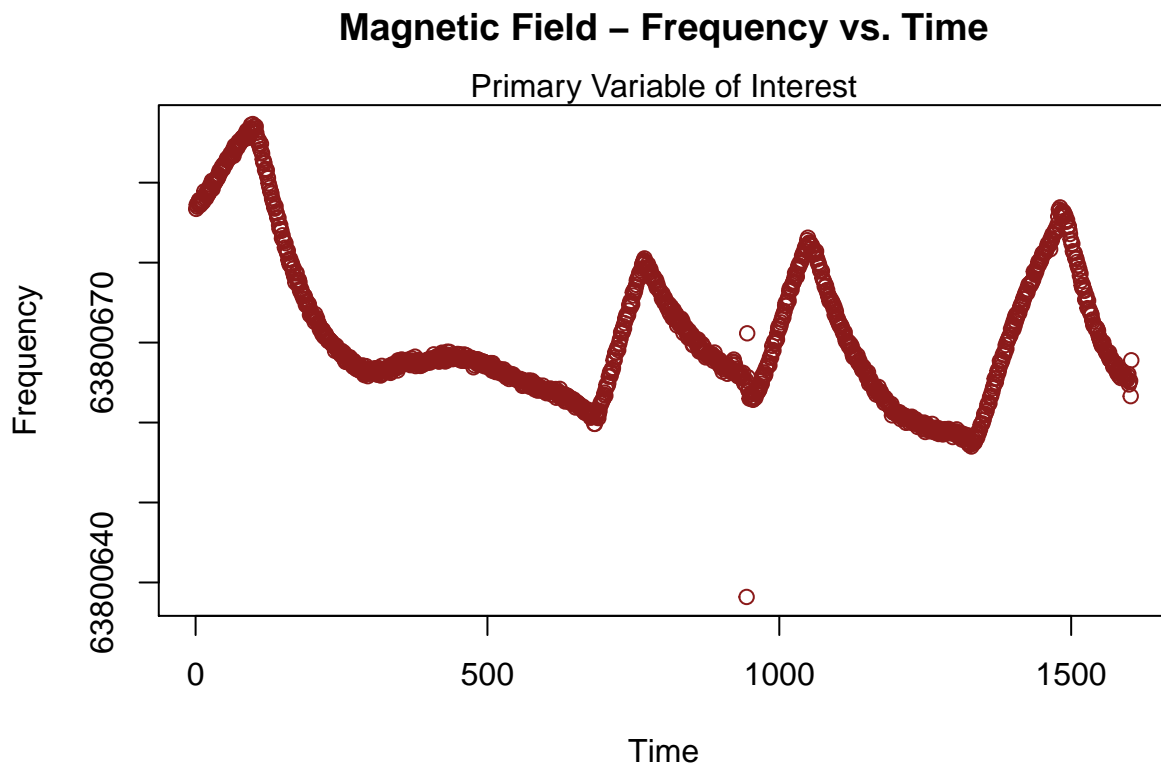
```
##      hour      minute      freq      int_pressure
## Min.   : 0.00   Min.   : 1.00   Min.   :63800638   Min.   :0.000
## 1st Qu.: 6.00   1st Qu.:15.00   1st Qu.:63800665   1st Qu.:2.969
## Median :13.00   Median :30.00   Median :63800668   Median :3.006
## Mean   :12.86   Mean   :30.27   Mean   :63800671   Mean   :3.011
## 3rd Qu.:20.00   3rd Qu.:45.00   3rd Qu.:63800676   3rd Qu.:3.055
## Max.   :26.00   Max.   :60.00   Max.   :63800697   Max.   :3.162
##
##   atm_pressure   tot_pressure      tesla
## Min.   :14.63   Min.   :17.60   Min.   :1.499
## 1st Qu.:14.66   1st Qu.:17.63   1st Qu.:1.499
```

```
## Median :14.67   Median :17.67   Median :1.499
## Mean   :14.66   Mean   :17.68   Mean   :1.499
## 3rd Qu.:14.68   3rd Qu.:17.72   3rd Qu.:1.499
## Max.   :14.69   Max.   :17.80   Max.   :1.499
##                                     NA's   :3
```

Remaking Eric's Plots

Primary Plot First

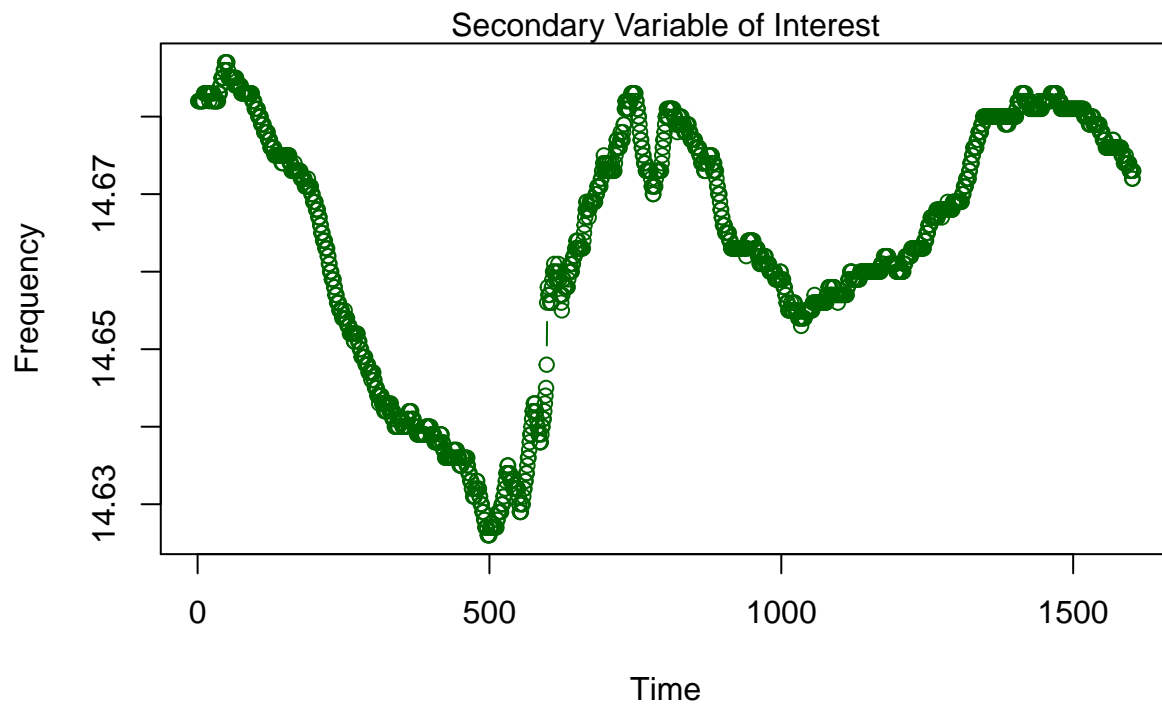
```
plot(mri$freq, col="firebrick4"
     , main="Magnetic Field - Frequency vs. Time"
     , xlab="Time"
     , ylab="Frequency"
     )
mtext("Primary Variable of Interest")
```



Covariate Plots Second

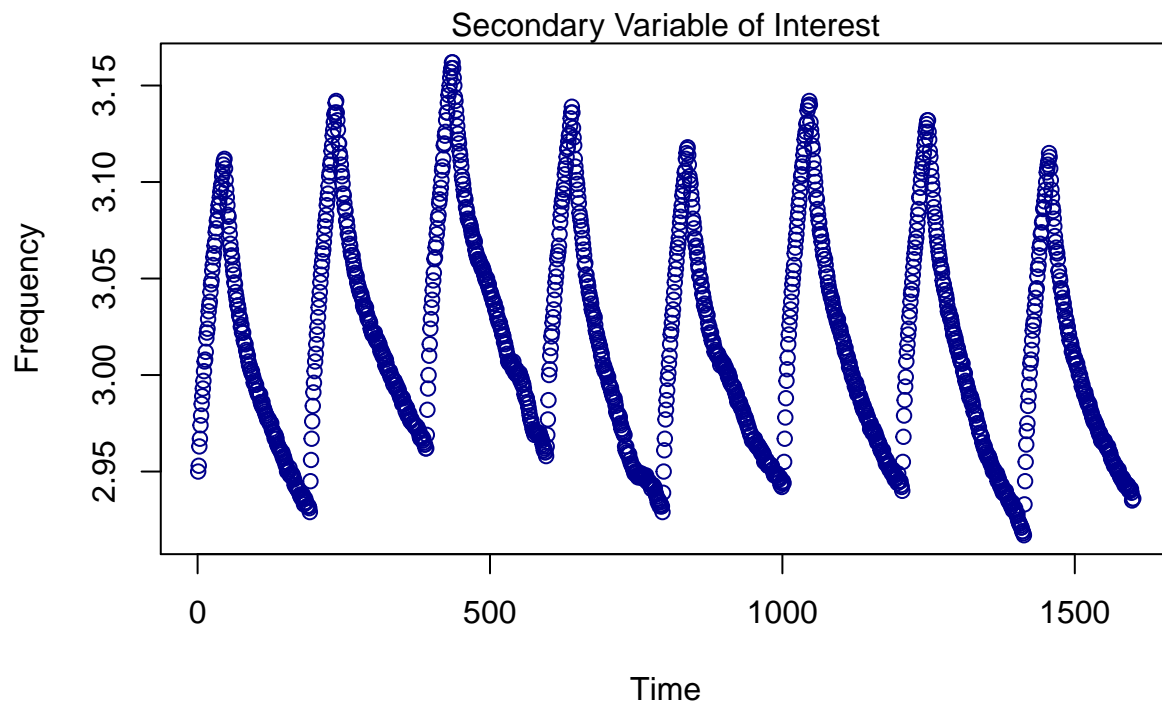
```
plot(mri$atm_pressure, col="darkgreen", type="b"
     , main="Atmospheric Pressure vs. Time"
     , xlab="Time"
     , ylab="Frequency"
     )
mtext("Secondary Variable of Interest")
```

Atmospheric Pressure vs. Time



```
plot(mri$int_pressure[1:1600], col="darkblue", type="b"  
     , main="Controlled Internal Pressure vs. Time"  
     , xlab="Time"  
     , ylab="Frequency"  
     )  
mtext("Secondary Variable of Interest")
```

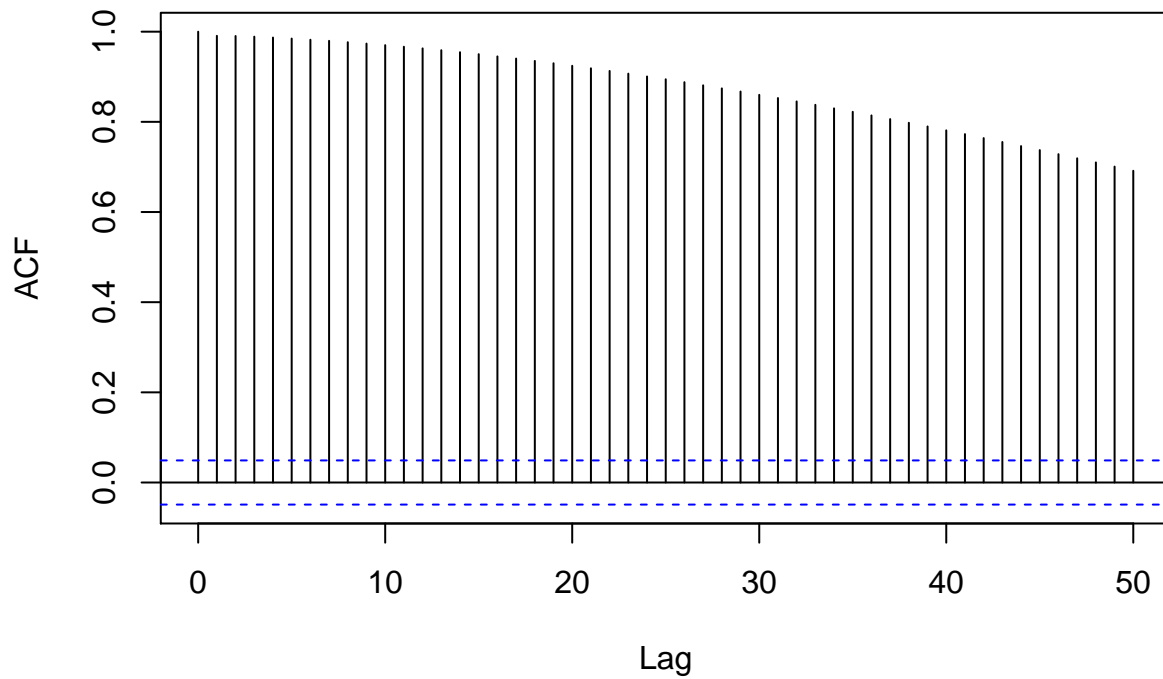
Controlled Internal Pressure vs. Time



Empirical Autocorrelation Functions

```
freq_acf <- acf(mri$freq, lag.max=50
               , main="ACF for Magnetic Frequency"
               )
```

ACF for Magnetic Frequency

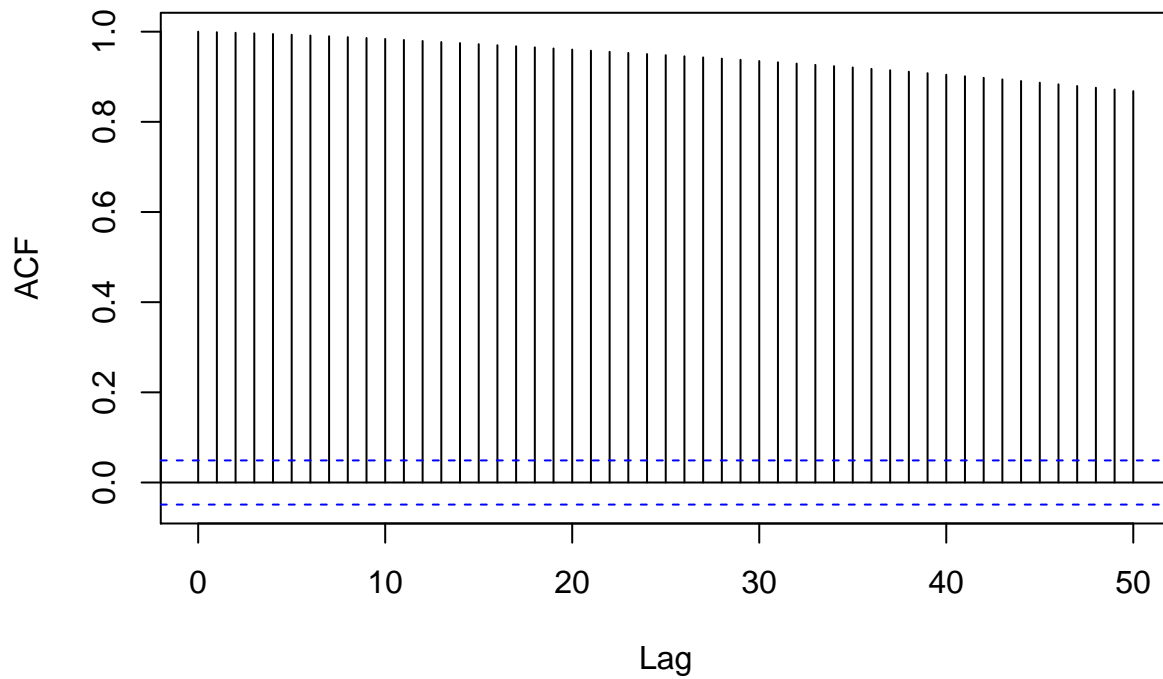


So our main variable has a LOT of serial correlation and is far from stationary. This could be because of time iteration. Consider aggregation to every five minutes as Eric mentioned.

ACF for Secondary Variables

```
acf(mri$atm_pressure, lag.max=50
    , main="ACF for Atmospheric Pressure"
    )
```

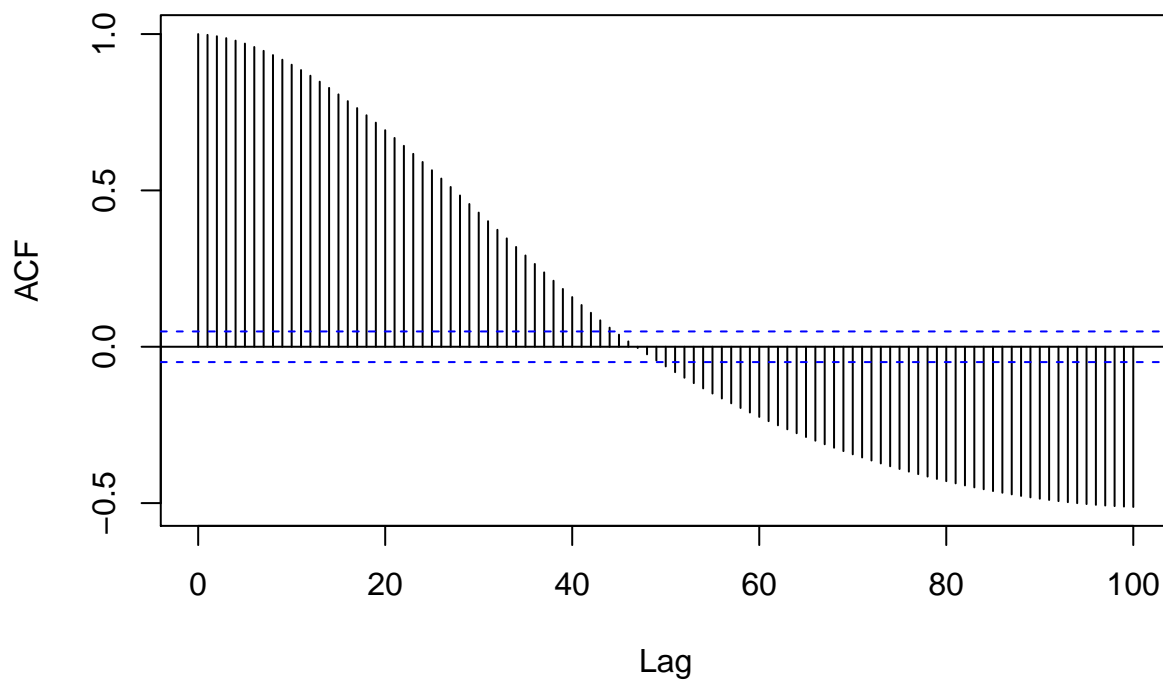
ACF for Atmospheric Pressure



Same story as above.

```
acf(mri$int_pressure[1:1600], lag.max=100
    , main="ACF - Internal Pressure"
    )
```

ACF – Internal Pressure



ACF for Internal Pressure variable is interesting - reflects the cyclic nature of series.

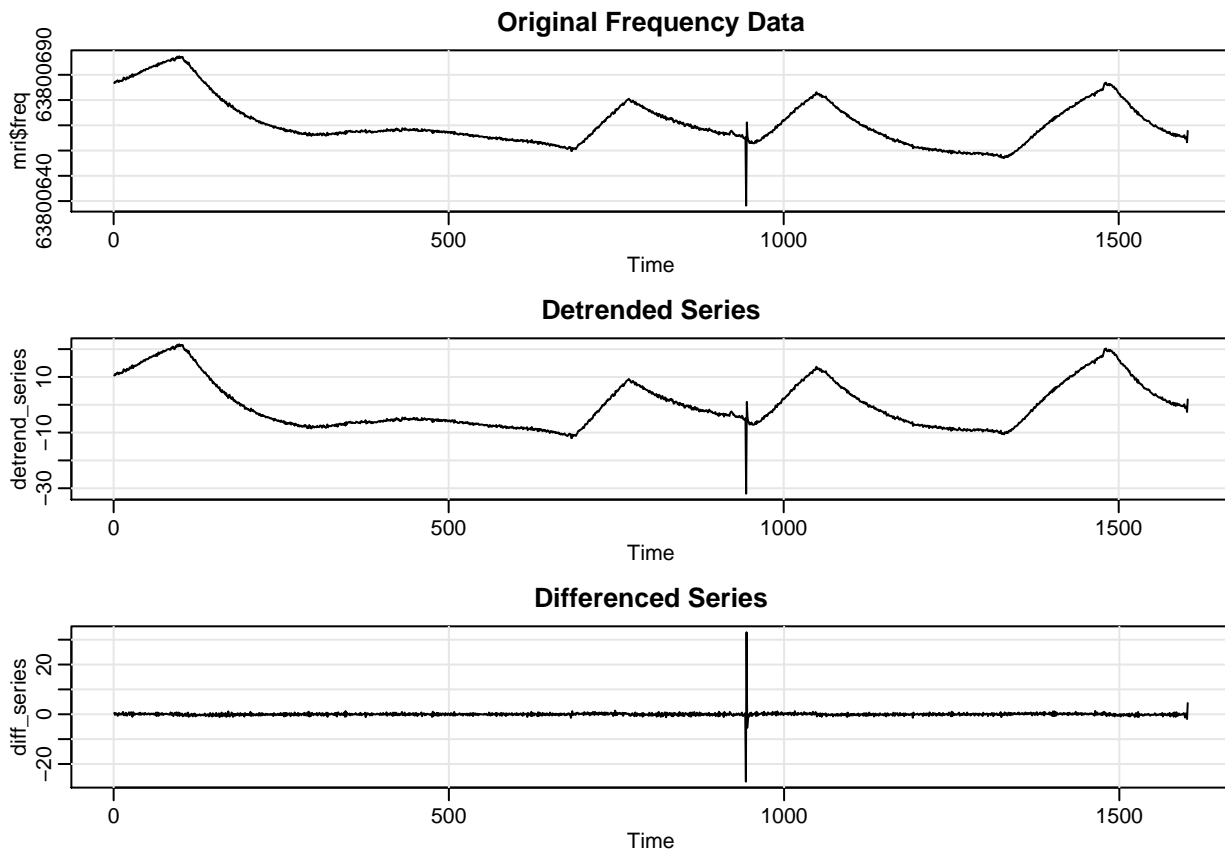
Primary Series Manipulations

Detrending + First Differences

```
par(mfrow=c(3,1))
# plot 1
astsa::tsplot(mri$freq, main="Original Frequency Data")

# plot 2
detrend_fit <- lm(mri$freq ~ time(mri$freq))
detrend_series <- resid(detrend_fit)
astsa::tsplot(detrend_series, main="Detrended Series")

# plot 3
diff_series <- diff(mri$freq)
astsa::tsplot(diff_series, main="Differenced Series")
```



The strange point in the middle may be influencing the images. It may be important, but let's take it out and try again.

```
print(length(mri$freq))
```

```
## [1] 1603
```

```

# take away minimum value
mri_freq2 <- mri$freq[-(which.min(mri$freq))]
# make sure that worked
print(length(mri_freq2))

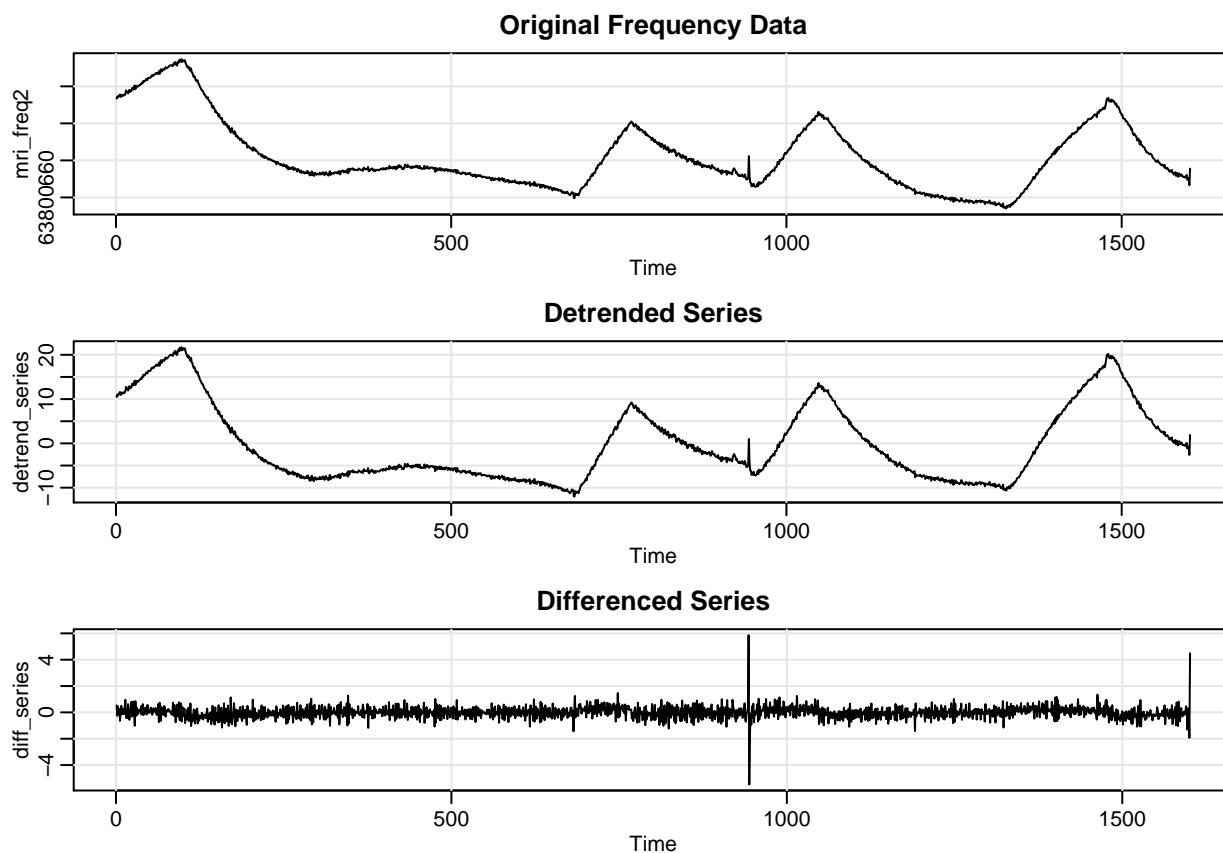
## [1] 1602

par(mfrow=c(3,1))
# plot 1
astsa::tsplot(mri_freq2, main="Original Frequency Data")

# plot 2
detrend_fit <- lm(mri_freq2 ~ time(mri_freq2))
detrend_series <- resid(detrend_fit)
astsa::tsplot(detrend_series, main="Detrended Series")

# plot 3
diff_series <- diff(mri_freq2)
astsa::tsplot(diff_series, main="Differenced Series")

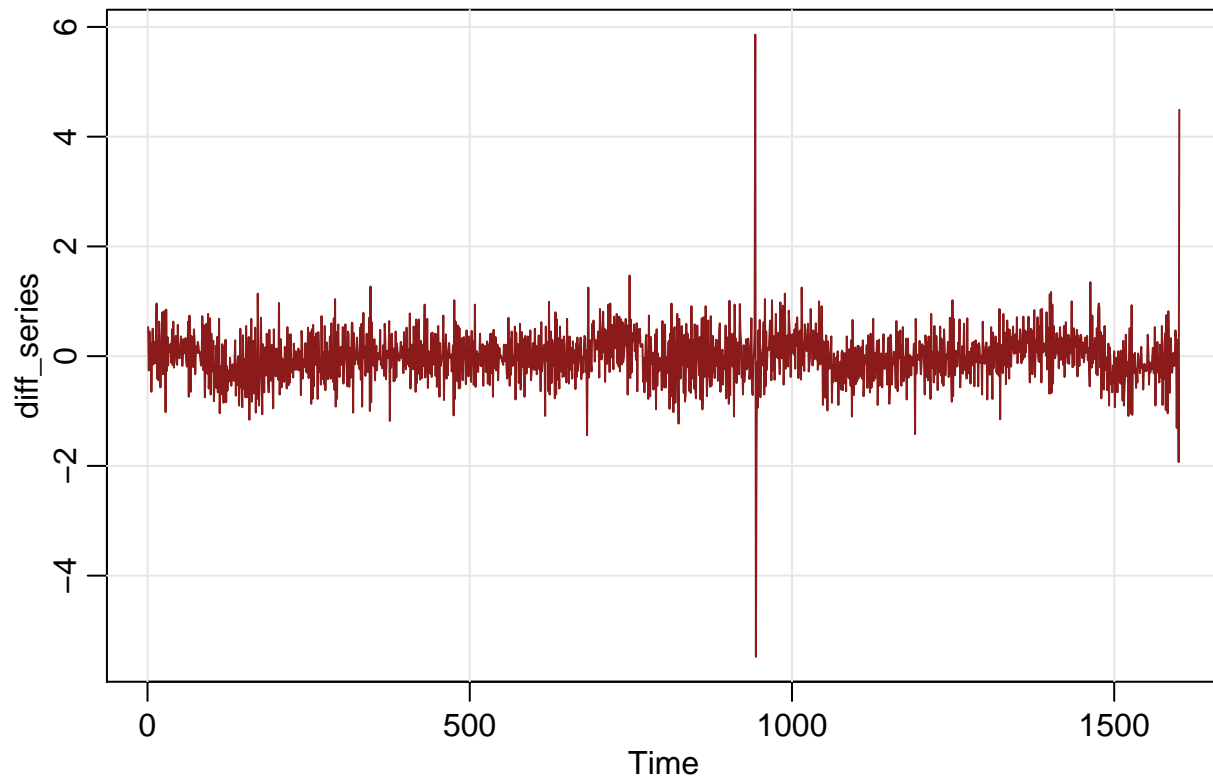
```



That's somewhat better. Detrending did nothing but change the axis. Differencing looks stationary?

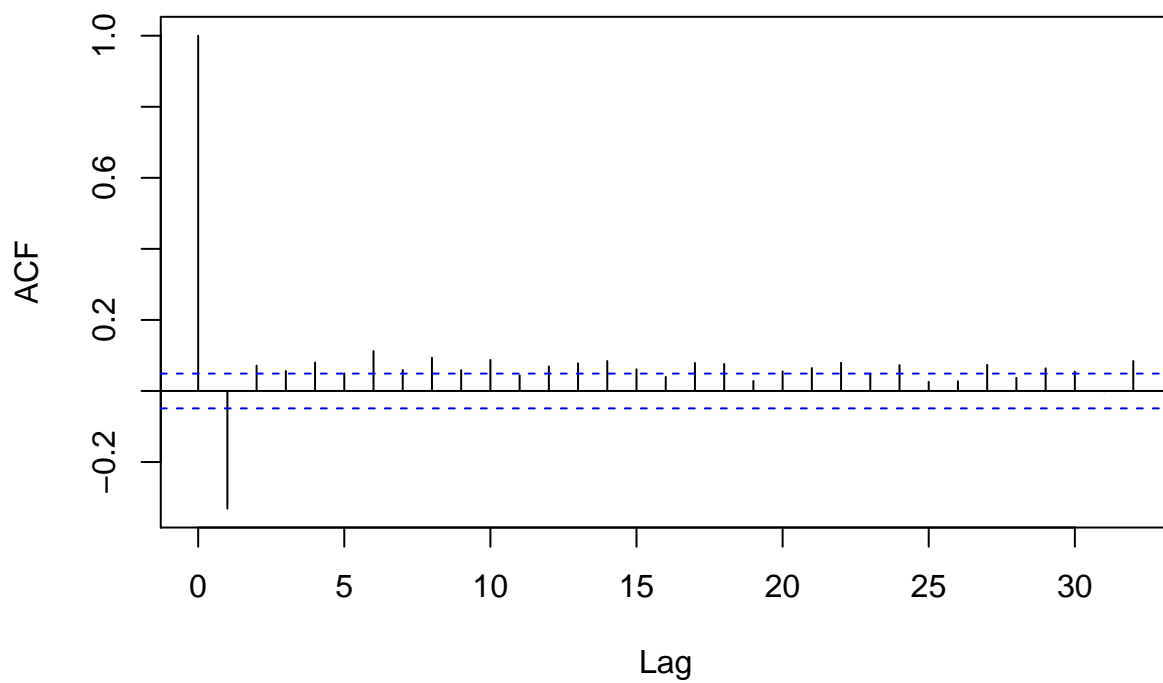
```
astsa::tsplot(diff_series, main="Differenced Series", col="firebrick4")
```


Differenced Series



```
acf(diff_series, main="ACF for Differenced Series")
```

ACF for Differenced Series



Hard to see, but the ACF at a lot of these lag values is significant. We are, however, much closer to stationarity

than we were.

Formal Stationarity Check

```
Box.test(mri$freq, type="Ljung-Box")

##
## Box-Ljung test
##
## data: mri$freq
## X-squared = 1576.8, df = 1, p-value < 2.2e-16
```

```
Box.test(detrend_series, type="Ljung-Box")

##
## Box-Ljung test
##
## data: detrend_series
## X-squared = 1598.1, df = 1, p-value < 2.2e-16
```

```
Box.test(diff_series, type="Ljung-Box")

##
## Box-Ljung test
##
## data: diff_series
## X-squared = 175.89, df = 1, p-value < 2.2e-16
```

My understanding of this procedure is that we are looking for a high p-value. We are consistently rejecting the null that the data are independently distributed (https://en.wikipedia.org/wiki/Ljung-T1\textendashBox_test).

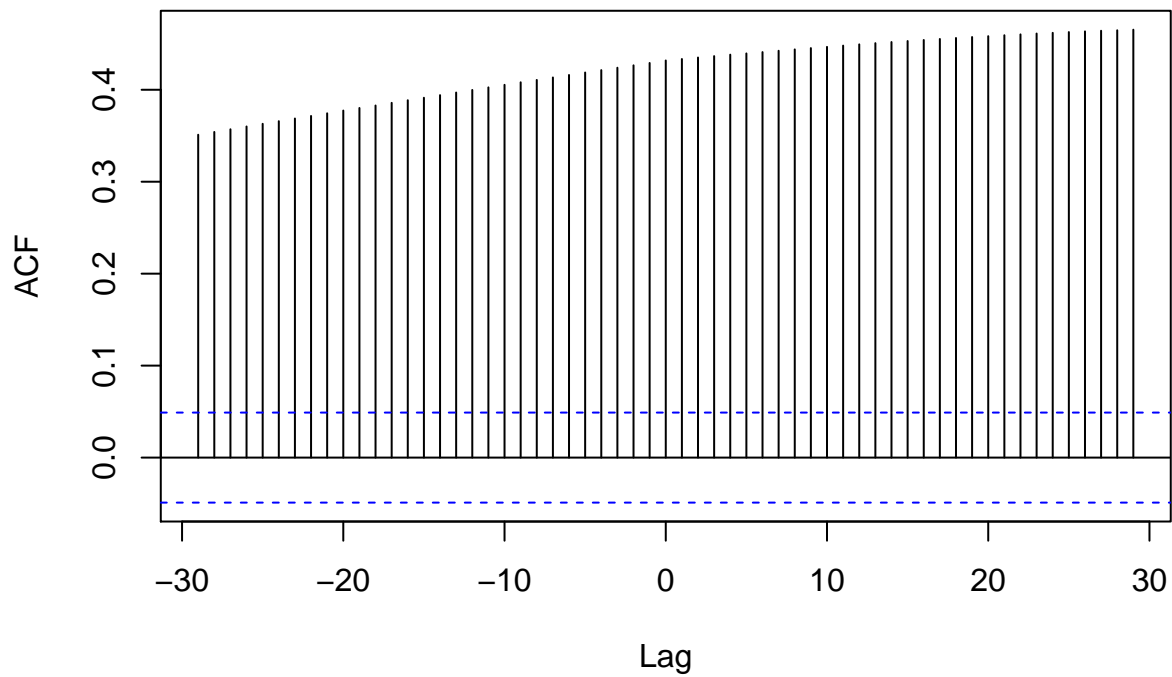
```
x <- rnorm(500, 0 ,1 )
Box.test(x, type="Ljung-Box")

##
## Box-Ljung test
##
## data: x
## X-squared = 1.3717, df = 1, p-value = 0.2415
```

Cross-Correlation Functions

```
ccf(mri$freq, mri$atm_pressure
    , main="CCF between Magnetic Frequency and Atmospheric Pressure"
    )
```

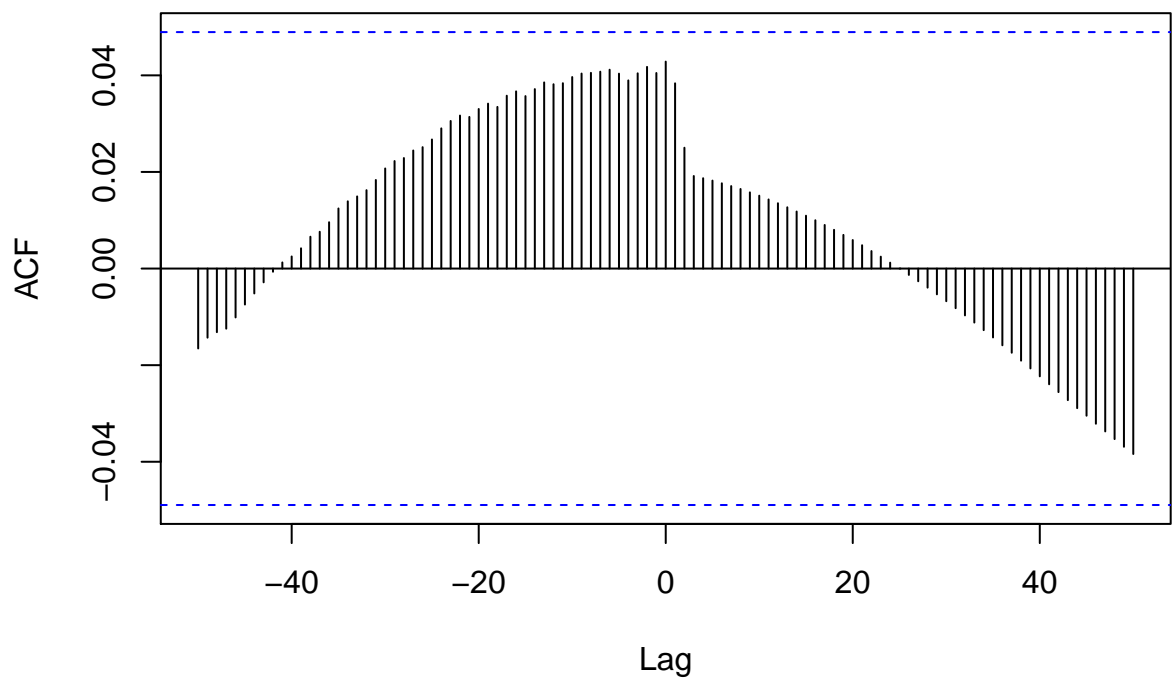
CCF between Magnetic Frequency and Atmospheric Pressure



Yikes.

```
ccf(mri$freq, mri$int_pressure, lag.max=50
    , main="CCF between Magnetic Frequency and Internal Pressure"
    )
```

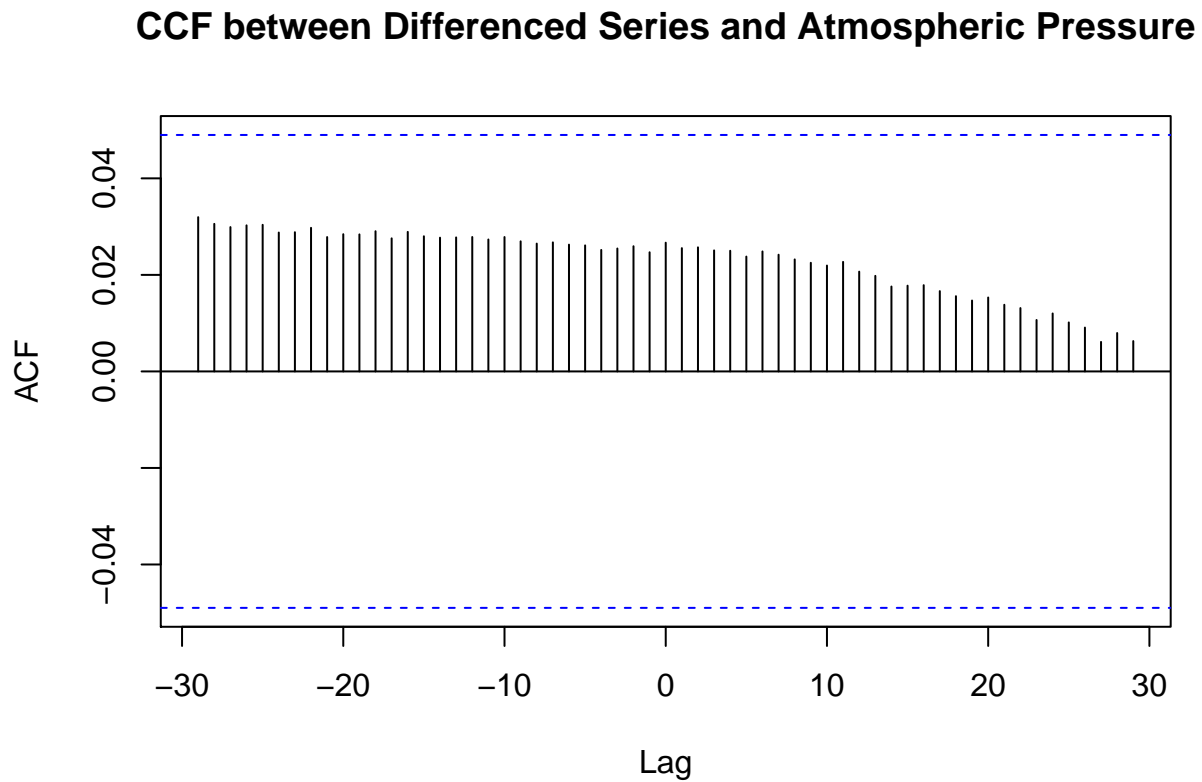
CCF between Magnetic Frequency and Internal Pressure



Also yikes, but less so.

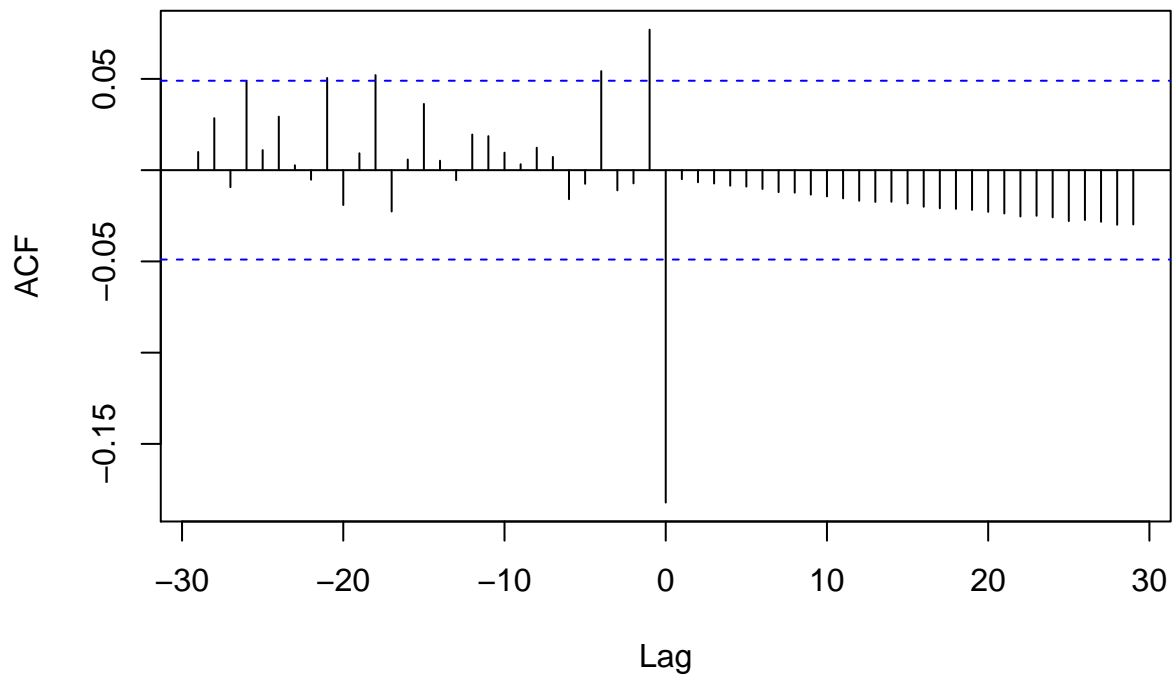
Let's swap out the main series for the differenced one.

```
ccf(diff_series, mri$atm_pressure
    , main="CCF between Differenced Series and Atmospheric Pressure"
    )
```



```
ccf(diff_series, mri$int_pressure
    , main="CCF between Differenced Series and Internal Pressure"
    )
```

CCF between Differenced Series and Internal Pressure



This looks promising.